

Fostering Open Access at CUL by the use of Open Access Repositories (OARs)

by
Ross Atkinson and Marcy Rosenkrantz

Introduction

Libraries in general and CUL in particular, have always been committed to providing free, open and un-impeded access to information. But as the cost of peer-reviewed scholarly literature, especially in the sciences, continues to increase we find ourselves being forced to make difficult decisions about how to allocate our limited resources. How do we strive to meet this commitment in the face of declining budgets and increasing costs? Several models of open-access publishing have been suggested, and task forces have been convened and have issued reports regarding those models. (see for example <http://techreports.library.cornell.edu:8081/Dienst/UI/1.0/Display/cul.lib/2004-3>) Ultimately we come to the same conclusion—we continue to make the *access* to scholarly information and literature free to our users—but providing that access has always and will continue to come at a cost to the library.

One way to help contain the costs of open access is to encourage scholars to publish in the growing number of open access journals rather than in the journals published by large for-profit companies. The former are inexpensive when compared to the latter. Another is to encourage pre- and post-publication of scholarly information in an open access repository or OAR. An OAR (in the specialized sense we are using the word here) is a digital container into which scholars can deposit digital objects, to ensure that such objects are easily accessible and are maintained for the longer term. The OAR, which CUL is committed to creating, is one that makes the objects it contains freely accessible to users of the Internet, and is designed so that its contents can be easily retrieved by standard search engines such as Google. Most such repositories of this type are maintained by individual research libraries.

All kinds of materials in any digital format (text, image, audio, video) can be added to the OAR. The OAR may contain formal publications (by which we mean publications that have passed some form of peer review), such as digital copies of journal articles or books. In such cases, the scholar adding the publication will need to have obtained agreement from the original publisher for copyright purposes. (Someday repositories may have high-level editorial boards themselves, so that repositories could serve as original publishers—rather than as containers of copies of items published elsewhere—but we are clearly still some distance from that eventuality.)

The OAR can also include all manner of informal (not peer reviewed) publications, such as presentations, working papers, data sets, or course lecture notes. Several institutions, including Cornell, provide graduate students with the option of adding their completed dissertations to the OAR. While the OAR is intended to be openly accessible, the Library

can conceal particular objects for a limited period of time, if the scholar plans on using the material for publication. The Library will accept materials for the OAR, however, only on the condition that they will be made openly accessible by a date upon which the scholar and the Library have agreed.

There are presently two common types of OARs. The most common are institutional repositories, such as the current DSpace implementation (<http://dspace.library.cornell.edu/>). These are intended for use by scholars (and, at Cornell, by students and staff) of the institution. The contents of most of these repositories are divided into communities, each one of which represents a particular community of interest such as an academic department or program. This provides each department or program with the ability to customize access and to create a collection policy that best meets the needs of those particular scholars.

The other main type is the disciplinary repository. This type, while currently far less prevalent than the institutional repository, is likely to become much more heavily used in the future. The disciplinary repository should ultimately become the standard, designated channel for the discipline. It should be the place the scholar goes to learn of the latest developments in the discipline, and also the place the scholar deposits his or her own publications. The most successful disciplinary repository at this time is arXiv (<http://www.arxiv.org/>), which includes papers on physics, computer science, mathematics, non-linear science, and quantitative biology. CUL maintains the arXiv and makes it freely and openly accessible to the world, on the assumption that other institutions will eventually develop and maintain similar repositories for other disciplines, to which Cornell scholars and students will have free and open access. It is also possible that institutional repositories may eventually serve primarily as the conduits to disciplinary repositories; a scholar wishing to add an item would simply put it into the central library repository at his or her institution; the library could then see that any necessary format adjustments are made, and the publication would then be transferred to the disciplinary repository located elsewhere. A separate file of all publications sent to disciplinary repositories might also be maintained in the institutional repository for preservation purposes.

The long-term goal of sophisticated, customized disciplinary repositories—toward which such services as the CUL OAR represent the first steps—is easy and open access to the core literature of each discipline for anyone who is interested in it. Repositories will also ensure much more effective access to informal (as well as formal) publications than has ever been possible before. Creating and maintaining such repositories, and encouraging scholars in all disciplines to make effective use of them, will likely become one of the central responsibilities of research libraries. The Cornell Faculty Senate and the Computer Science Department have put forth a resolution and policy, respectively, encouraging open access, and the use of an OAR (<http://www.library.cornell.edu/scholarlycomm/resolution.html>, <http://www.cs.cornell.edu/home/vavasis/arxiv.html>).

As a result of the growing interest in OARs here and elsewhere, several CUL staff have been asked to give talks about the systems employed at CU (primarily DSpace and arXiv). At these presentations we usually cover the what, why, how, and where of open access. We've tried to summarize the answers to some of these questions in the Frequently Asked Questions that follow.

Frequently Asked Questions About Open Access Repositories

Why should I use an OAR?

CUL is committed to the concept of open access for many reasons. It opens scholarly literature to the world and promotes the open exchange of ideas. Without it only the people or institutions that can afford to pay will have access. That would likely exclude third world countries from getting access to information that would help them grow. Indeed, physics scholars in the Former Soviet Union have claimed that without arXiv, they would have little if any access to the current physics literature.

What is DSpace?

DSpace is both the name of the software system that underlies the digital (institutional) repository of the same name. Several institutions using the DSpace software have re-named their repositories to avoid confusion. (see for example <https://tspace.library.utoronto.ca/> or <https://urresearch.rochester.edu/index.jsp/>)

I understand there are several digital repository software systems. Why are we using DSpace instead of one of them?

There are several software systems that could be used for a digital repository

- DPubS, which was developed initially in Cornell's Computer Science Department and further, developed in CUL. It is the software system for Project Euclid and several digital collections, as well as <http://techreports.library.cornell.edu> used primarily by the Computer Science Department,
- E-Prints, used primarily in Europe but also at Cal Tech, and
- BePrints, a proprietary software system used as for electronic publishing and as a repository at the California Digital Library project. (BePrints is also used as the repository system at Cornell's ILR School.

We chose DSpace because it is free, open source software and has a large community using it and contributing to the code base. As a result, its features and functionality are growing relatively rapidly.

How are issues related to intellectual property handled in DSpace?

Authors who submit work to DSpace are asked to agree to a simple license agreement that grants CUL non-exclusive rights to distribute, reproduce, and translate the work. See a copy of the agreement at the end of this article. University Counsel has vetted the license agreement.

If I submit my work to DSpace and later decide to submit it to a journal or a book publisher, won't publishers reject it because it has already been published? And even if the publisher accepts it, can I put a post-print online, too?

Although some publishers are uncomfortable accepting work that has appeared in a repository or your own web site, and some are even downright hostile to the idea, the number of them that will accept work that has already appeared online is growing. You can determine the policies of many journal publishers regarding prior publication online at the [Sherpa web site](#). Look for the "green" publishers. They are the ones with few if any restrictions concerning online publication of pre-prints and post-prints. You'll find

most publishers will allow you to keep a preprint in your institution's repository or on your own web site. Many will let you post your own copy of the "as published work" as long as you refer to the published work appropriately and link to it if it is online at the publisher's site.

Most publishers make me sign a copyright transfer agreement, which means that I no longer own the rights to the work. Do I have to transfer all my rights? And if I do, wouldn't that force me to take my work out of DSpace?

Although they don't advertise it, most publishers will allow you to reserve some rights for yourself. SPARC has made the [author's addendum](#) available. We encourage you to add it to any copyright transfer agreement you sign. CUL also maintains a web site for [copyright management](#) that you will find helpful.

Can I take an object out of DSpace once I submit it?

Let's answer that with another question. If you submit something to a journal, will they allow you to remove it later? The answer is probably not. So we don't want to remove objects from DSpace unless required to do so for a legal reason, e.g. you didn't have the right to submit it in the first place. Even under such a circumstance there would be a "tombstone" indicating there had been an item in DSpace but it was removed.

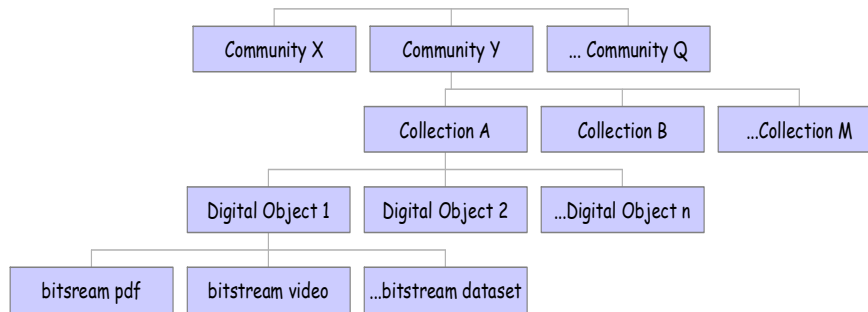
I don't want the world to see my work. Can I select who can download my DSpace object?

DSpace has a provision for closed collections, but we'd prefer not to allow such submissions. We are trying to promote open access, not discourage it, after all. That said, we do have a closed collection for theses and dissertations. Unfortunately, too many students are submitting to that collection, under the mistaken impression that all publishers will reject work previously posted online. See the response to the earlier question regarding prior publication.

You've been talking a lot about objects in DSpace. What do you mean?

DSpace is organized hierarchically. It has communities, sub-communities, collections, digital objects, and files, or bitstreams (a collection of 1's and 0's). A community is some organizational unit, like a department, college, or school with some common interest or purpose. A community can spawn sub-communities (e.g. a college might want each department to be a sub-community of the college community). And a sub-community could divide itself into one or more sub-sub-communities. But let's not push that too far. Within any community level there can be one or more collections of digital objects. Each object can be comprised of files that are related to each other. For instance, say you have a collection of books. Each book in the collection is an object and the chapters of the book are the files that make up the object. Perhaps the following diagram will help.

DSpace's Organizational Structure



You've been using the term digital repository. I've also heard about institutional repositories. What's the difference?

Colleagues at other institutions have found that when they talk to faculty about institutional repositories, by and large there is great resistance to their use. "It's my work, not the institution's," they claim. And rightfully so. But if you call it a digital repository you remove the concept of institutional ownership and faculty are more accepting of the idea. The two are pretty much the same thing.

What's the difference between a digital repository and arXiv?

[ArXiv](#) is a digital repository for scholarly communication in physics, computer science, mathematics, non-linear sciences, and quantitative biology. It's managed and administered by CUL but it is submitted to and used by scholars world-wide, not just faculty, staff, students, and researchers at Cornell. Like arXiv, we want the objects in our OAR to be used, read, and downloaded by people all over the world, but at the present time submission to DSpace is limited to the Cornell community.

How can I be sure that if I submit an object to DSpace or any OAR it will be there forever?

We can't speak for OARs at other institutions or those used elsewhere at Cornell, but we can assure you that we consider DSpace to be an integral part of any digital preservation system we put in place. And the implementation of an OAR and the development of a digital preservation system are CUL priorities (priority teams 9 and 2, respectively). Certainly, it's safer than maintaining your own web site.

NON-EXCLUSIVE DISTRIBUTION LICENSE

By signing and submitting this license, you (the author(s) or copyright owner) grants to the Cornell University Library (CUL) the non-exclusive right to reproduce, translate (as defined below), and/or distribute your submission (including the abstract) worldwide in print and electronic format and in any medium, including but not limited to audio or video.

You agree that CUL may, without changing the content, translate the submission to any medium or format for the purpose of preservation.

You also agree that CUL may keep more than one copy of this submission for purposes of security, back-up and preservation.

You represent that the submission is your original work, and that you have the right to grant the rights contained in this license. You also represent that your submission does not, to the best of your knowledge, infringe upon anyone's copyright.

If the submission contains material for which you do not hold copyright, you represent that you have obtained the unrestricted permission of the copyright owner to grant CUL the rights required by this license, and that such third-party owned material is clearly identified and acknowledged within the text or content of the submission.

IF THE SUBMISSION IS BASED UPON WORK THAT HAS BEEN SPONSORED OR SUPPORTED BY AN AGENCY OR ORGANIZATION OTHER THAN CUL, YOU REPRESENT THAT YOU HAVE FULFILLED ANY RIGHT OF REVIEW OR OTHER OBLIGATIONS REQUIRED BY SUCH CONTRACT OR AGREEMENT.

CUL will clearly identify your name(s) as the author(s) or owner(s) of the submission, and will not make any alteration, other than as allowed by this license, to your submission.